

Tilburg University

Asymptotic power of the integrated conditional moment test against global and large local alternatives

Ploberger, W.; Bierens, H.J.

Publication date:
1995

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Ploberger, W., & Bierens, H. J. (1995). *Asymptotic power of the integrated conditional moment test against global and large local alternatives*. (CentER Discussion Paper; Vol. 1995-122). Unknown Publisher.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Asymptotic Power of the Integrated Conditional Moment Test Against Global and Large Local Alternatives.

Werner PLOBERGER ¹

University of St. Andrews, Scotland

Herman J. BIERENS

Southern Methodist University, Dallas, U.S.A, and

Tilburg University, the Netherlands.

In this paper we study further the asymptotic power properties of the integrated conditional moment (*ICM*) test of Bierens (1982) and Bierens and Ploberger (1994). First, we establish the relation between consistency against global alternatives and nontrivial local power, using the concept of potential consistency. Moreover, we study the asymptotic power of the test under a class of "large" local alternatives that shrink to the null at rate $O_p(c/\sqrt{n})$, where n is the sample size and c is a large positive constant. We show that the local asymptotic power of the ICM test can be made arbitrarily close to 1 by choosing this constant c sufficiently large, where the rate of convergence is essentially independent of the instruments. Furthermore, we compare the asymptotic power of the ICM test against these large local alternatives with the asymptotic power of the parametric t -test. The asymptotic power function of the t -test under large local alternatives approaches 1 at the same rate as the consistent ICM test for $c \rightarrow \infty$ only if the local alternative is correctly specified up to a constant c . In all other cases the ICM test is asymptotically more powerful.

September 1, 1995

¹ Most of the research in this paper was done while the first author was affiliated with the Technische Universität Wien, Austria. The financial support from the "Ausseninstitut der Technische Universität Wien" is gratefully acknowledged.

1. INTRODUCTION

The Integrated Conditional Moment (ICM) test of functional form was proposed by Bierens (1982). Bierens's test is the first test proposed in the literature that is consistent against all deviations from the null hypothesis that a given regression model represents the conditional expectation of the dependent variable relative to the vector of regressors. The ICM test is based on an integral over a squared weighted mean of the least squares residuals, where the weights involved are random functions on a subset of a Euclidean space. Bierens (1982) showed that these weight functions can be chosen such that under the alternative of misspecification of functional form this weighted mean of least squares residuals converges in probability to a continuous function that is not everywhere equal to zero, so that the ICM test is consistent. In later work, Bierens (1984, 1987, 1990), Bierens and Hartog (1988), De Jong (1995), De Jong and Bierens (1994), Lewbel (1992) and Stinchcombe and White (1991) employed the latter result in designing more general and/or alternative consistent tests of functional form. Also White's (1989) neural network test belongs to this class of consistent tests of functional form. In section 3 we address some issues regarding the consistency of the ICM test, in relation with the nontrivial local power.

In Bierens and Ploberger (1994) we have analyzed the local power of the ICM test. In particular, we showed that the ICM test has nontrivial asymptotic power against \sqrt{n} -local alternatives. These results are briefly reviewed in section 2. Note that, next to the conditional moment tests based on extensions of Bierens' (1982) approach, there is also a class of alternative consistent tests of functional form based on comparison of parametric and (semi-) nonparametric models. See, e.g., Wooldridge (1992), Yatchew (1992), Gozalo (1993), Hardle and Mammen (1993), Horowitz and Hardle (1992), Fan and Li (1992a,b,c), Hong and White (1991), White and Hong (1993), and Zheng (1993), among others. However, it seems that these type of tests have only nontrivial power in all local directions for local alternatives that shrink to the null at a slower rate than $1/\sqrt{n}$. Therefore, these tests are less powerful than the ICM test.

In Bierens and Ploberger (1994) we also showed that the ICM test is admissible, in the sense that asymptotically there does not exist a uniformly more powerful test. Moreover, we solved the size problem that hampered the application of original ICM test of Bierens (1982) by deriving sharp upperbounds of the critical values of the ICM test.

In this paper we study further the asymptotic power properties of the integrated conditional moment (*ICM*) test of Bierens (1982) and Bierens and Ploberger (1994). First, in section 2, we briefly review the asymptotic theory of the ICM test under local alternatives. In section 3, we establish the relation between consistency against global alternatives and nontrivial local power, using the concept of potential consistency. Moreover, we study the asymptotic power of the test under a class of "large" local alternatives. These large local alternatives take the form of a semi-parametric augmented regression model such that the difference between this augmented regression model and the null model is orthogonal to the derivatives of the regression function and vanishes to zero at order $O_p(c/\sqrt{n})$, where c is a large positive constant. Finally, in section 4 we show that under regularity conditions the local asymptotic power of the ICM test can be made arbitrarily close to 1 by choosing this constant c sufficiently large. The rate of convergence to 1 of the power function is exponential and independent of the instruments and the form of the alternatives, provided the test is consistent. Moreover, we compare the asymptotic power of the ICM test against these large local alternatives with the asymptotic power of the parametric t -test. It appears that in general the asymptotic power function of the consistent ICM test converge to 1 for $c \rightarrow \infty$ at a faster exponential rate than the t -test. Only if the t -test is conducted on the basis of the local alternative itself, thus assuming that the alternative is completely parametric and all the variables involved are observable, the rate of convergence to 1 of the asymptotic power functions of the consistent ICM test and the t -test is the same. This is surprising. The common intuition is that a consistent test spreads its power thinly over all possible alternatives, and that therefore a test that is designed to have optimal power against a particular alternative is in general more powerful in a neighborhood of this alternative than a consistent test. The results in this paper refute this.

Throughout this paper we shall refer to Assumption A in Bierens and Ploberger (1994) as "Assumption A".

2. THE LIMITING DISTRIBUTION OF THE ICM TEST STATISTIC UNDER LOCAL ALTERNATIVES

The ICM test is a test for the correctness of a parametric regression model $y_t = f_t(\theta_0) + u_t$, where the y_t 's are the dependent variables, the $f_t(\cdot)$'s are the regression functions, depending on lagged dependent variables and/or exogenous variables x_t , θ_0 is an unknown parameter vector and u_t is a martingale difference error process. More precisely, the null hypothesis to be tested is:

$$(1) \quad H_0: y_t = f_t(\theta_0) + u_t, \theta_0 \in \Theta,$$

under the maintained hypothesis that Assumption A holds. The relevant parts of this assumption for the null hypothesis (1) are that the parameter space Θ is a compact subset of \mathbb{R}^n , with the true parameter vector θ_0 contained in its interior, the response function $f_t(\theta)$ is twice continuously differentiable on Θ , and u_t and $f_t(\theta)$ are measurable w.r.t. to the σ -algebra \mathcal{F}_t generated by x_t and $(y_{t-j}, x_{t-j}), j = 0, 1, 2, \dots$, and the crucial condition that the errors u_t form a martingale difference sequence, i.e., $E(u_t | \mathcal{F}_t) = 0$ a.s.²

In order to unify the time series and cross-section cases, we need to augment Assumption A as follows:

ASSUMPTION B: $f_t(\theta)$ is measurable w.r.t. to the σ -algebra \mathfrak{G}_t , and $\mathcal{F}_t = \bigvee_{j \leq t} \mathfrak{G}_j$.

The latter means that \mathcal{F}_t is the minimum σ -algebra containing \mathfrak{G}_j for $j \leq t$. For example, in the independent linear regression case $f_t(\theta) = \theta_1 + \theta_2^T x_t$, the σ -algebra \mathfrak{G}_t is the σ -algebra generated by the vector x_t of regressors, whereas in the time series case $\mathfrak{G}_t = \mathcal{F}_t$. The reason for this distinction between \mathfrak{G}_t and \mathcal{F}_t is first that now in either case $\{u_t, \mathcal{F}_t\}$ is a martingale difference

² In Bierens and Ploberger (1994) we have assumed that $f_t(\theta)$ is measurable w.r.t. to the σ -algebra \mathcal{F}_{t-1} , where the latter has the same meaning as the present \mathcal{F}_t .

process³, and second that we can confine the alternatives to those that are measurable w.r.t. \mathfrak{G}_t .

Assuming that y_t and $f_t(\cdot)$ are observable for $t = 1, \dots, n$, the test statistic of the ICM tests is of the form

$$(2) \quad \hat{T} = \int \hat{z}(\xi)^2 d\mu(\xi), \quad \text{where } \hat{z}(\xi) = (1/\sqrt{n}) \sum_{t=1}^n (y_t - f_t(\hat{\theta})) w_t(\xi),$$

with $\hat{\theta}$ the least squares estimator of θ_0 , μ a probability measure on Ξ , and $w_t(\cdot)$ a sequence of real valued random weight functions on Ξ , where Ξ is a compact subset of a Euclidean space, such that $w_t(\xi)$ is measurable w.r.t. \mathfrak{G}_t . In Bierens and Ploberger (1994) we have analyzed the asymptotic power properties of the ICM test against local alternatives of the form:

$$(3) \quad H_1^L: y_{t,n} = f_t(\theta_0) + g_t/\sqrt{n} + u_t,$$

where g_t is measurable w.r.t. \mathfrak{G}_t . Note that the ICM test does not employ any a priori information on the g_t 's. Thus the local alternative (3) is actually a semi-parametric regression model, with $f_t(\theta)$ the parametric part and g_t the nonparametric part.

Under the local alternative (3) and some mild regularity conditions we can write the random function $\hat{z}(\xi)$ in (2) as

$$(4) \quad \hat{z}(\xi) = (1/\sqrt{n}) \sum_{t=1}^n u_t \phi_t(\xi) + (1/n) \sum_{t=1}^n g_t \phi_t(\xi) + o_p(1),$$

pointwise in ξ , where (again pointwise in ξ),

$$(5) \quad \begin{aligned} \phi_t(\xi) &= w_t(\xi) - \left[\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n (\partial/\partial \theta) f_j(\theta_0) w_j(\xi) \right] \\ &\times \left[\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n \left((\partial/\partial \theta^T) f_j(\theta_0) \right) \left((\partial/\partial \theta) f_j(\theta_0) \right) \right]^{-1} (\partial/\partial \theta^T) f_t(\theta_0). \end{aligned}$$

Then we have [cf. Bierens and Ploberger (1994)]:

³ Note that in the independent case $\{u_t, \mathfrak{G}_t\}$ is an independent zero mean process but in general not a martingale difference process, because \mathfrak{G}_t is not necessarily monotonic.

THEOREM 1: Under Assumptions A and B and H_1^L , $\hat{z} \Rightarrow z$, where z is a Gaussian process on Ξ with mean function

$$(6) \quad \eta(\xi) = \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n g_t \phi_t(\xi).$$

and covariance function

$$(7) \quad \Gamma(\xi_1, \xi_2) = \lim_{n \rightarrow \infty} (1/n) \sum_{t=1}^n E(u_t^2 \phi_t(\xi_1) \phi_t(\xi_2)).$$

Consequently, by the continuous mapping theorem, $\hat{T} \rightarrow T$ in distr., where $T = \int z^2(\xi) d\mu(\xi)$.

Moreover, denoting the eigenvalues and corresponding eigenfunctions of the covariance function Γ by λ_i and $\psi_i(\cdot)$, respectively, i.e.,

$$(8) \quad \int \Gamma(\xi_1, \xi_2) \psi_i(\xi_2) d\mu(\xi_2) = \lambda_i \psi_i(\xi_1), \quad i = 1, 2, \dots,$$

where $\lambda_i \geq 0$ and the $\psi_i(\cdot)$ are real valued such that

$$(9) \quad \int \psi_i(\xi) \psi_j(\xi) d\mu(\xi) = I(i = j),$$

with $I(\cdot)$ the indicator function, we can write $T = \sum_{i=1}^{\infty} (\eta_i + \varepsilon_i \sqrt{\lambda_i})^2$, where

$$(10) \quad \eta_i = \int \eta(\xi) \psi_i(\xi) d\mu(\xi)$$

and the ε_i are i.i.d. $N(0,1)$.

3. CONSISTENCY AND NONTRIVIAL LOCAL POWER

3.1. Potential consistency

As shown earlier by Bierens (1982, 1990), in the case of an i.i.d. data-generating process (y_t, x_t) , where the x_t 's are $k \times 1$ vectors of exogenous variables, and a null hypothesis of the form $P(E(y_t | \mathfrak{G}_t) = f_t(\theta_0)) = 1$ for some $\theta_0 \in \Theta$, where now $f_t(\theta) = f(x_t, \theta)$ with f a given function, we can choose the weight functions $w_t(\xi)$ and the probability measure μ such that the ICM test is consistent against all (*global*) deviations from the null hypothesis, i.e., if $P(E(y_t | \mathfrak{G}_t) = f_t(\theta)) < 1$ for all $\theta \in \Theta$, then $\text{plim}_{n \rightarrow \infty} \hat{T}/n > 0$. For example, in the i.i.d. case the choice $w_t(\xi) = \exp(\Phi(x_t)^T \xi)$ with Φ a bounded one-to-one mapping, and μ the uniform measure on a nonempty compact subset Ξ of the k -dimensional Euclidean space, yields a consistent ICM test. Of course, the ICM test is in general not consistent against \sqrt{n} -local alternatives of the form (3).

In the case where the data-generating process (y_t, x_t) is a vector time series process, with $f_t()$ depending on lagged values of y_t and x_t , consistency of the ICM test requires that the weight function $w_t()$ depends on *all* lagged values of y_t and x_t and an infinite-dimensional ξ . This case is considered by De Jong (1995). Therefore, if the weight function $w_t()$ depends only on a finite number of lagged values of y_t and x_t and a conformable finite-dimensional ξ , the ICM test cannot be consistent against all possible global alternatives, although the ICM test may still be consistent against a wide class of alternatives. In view of the results in Bierens (1982, 1984, 1990) we may characterize this class of global alternatives as follows, using the concept of potential consistency:

DEFINITION 1: Let \mathfrak{F}_t be the σ -algebra generated by $\{w_t(\xi), \xi \in \Xi\}$. The ICM test is potentially consistent if for any sequence a_t of random variables with finite second moments, and every t , $\int (E[a_t w_t(\xi)]^2) d\mu(\xi) = 0$ if and only if $E[a_t | \mathfrak{F}_t] = 0$ a.s.

Note that the potential consistency of the ICM test is only a matter of the right choice of the functional form of the weight function $w_t(\xi)$, given a set of conditioning variables, and the probability measure μ , and that this choice can be made independently of the data-generating process. The essence of this concept is that given a vector z_t of instruments generating \mathfrak{F}_t , we should choose the functional form of the weight function $w_t(\xi)$ and the measure μ such that if $P(E(a_t | \mathfrak{F}_t) = 0) < 1$ then the set

$$S = \{\xi \in \Xi : E(a_t w_t(\xi)) = 0\}$$

has positive μ measure. In particular, if the alternative is $y_t = f_t(\theta_0) + g_t + u_t$, where g_t is measurable \mathfrak{F}_t , and $P(E(g_t | \mathfrak{F}_t) = 0) < 1$, then the set S with $a_t = g_t + u_t$ has positive μ measure, so that this alternative is in principle identifiable. Of course, the practical identifiability depends on other condition as well, in particular stationarity. Anyhow, we can always choose the functional form of $w_t(\xi)$ and μ such that the ICM test is potentially consistent.

Under Assumption B the null hypothesis (1) is equivalent to the hypothesis $H_0: P[E(y_t | \mathfrak{F}_t) = f_t(\theta_0)] = 1$ for some $\theta_0 \in \Theta$, and the most general global alternative is that

the null is false: $H_1^G: P[E(y_t|\boldsymbol{\Theta}_t) = f_t(\boldsymbol{\theta})] < 1$ for all $\boldsymbol{\theta} \in \Theta$. This global alternative can also be written as

$$(11) \quad H_1^G: y_t = f_t(\boldsymbol{\theta}_0) + g_t + u_t,$$

where $f_t()$, g_t and u_t are the same as before and $\boldsymbol{\theta}_0$ is now the probability limit of the nonlinear least squares estimator $\hat{\boldsymbol{\theta}}$ under this alternative (Of course, $\boldsymbol{\theta}_0$ may be different under the null and the alternative).

In order to link the consistency property to the nontrivial local power property of the ICM test, and vice versa, we assume stationarity and some further regularity conditions:

ASSUMPTION C: $f_t(\boldsymbol{\theta})$, $(\partial/\partial\boldsymbol{\theta}^T)f_t(\boldsymbol{\theta})$, g_t , and $w_t(\xi)$ are stationary. Furthermore,

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n g_t w_t(\xi) &= E[g_1 w_1(\xi)], \\ \text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n w_t(\xi) (\partial/\partial\boldsymbol{\theta}^T)f_t(\boldsymbol{\theta}_0) &= E[w_1(\xi) (\partial/\partial\boldsymbol{\theta}^T)f_1(\boldsymbol{\theta}_0)], \end{aligned}$$

uniformly on Ξ , and

$$\text{plim}_{n \rightarrow \infty} (1/n) \sum_{t=1}^n g_t (\partial/\partial\boldsymbol{\theta}^T)f_t(\boldsymbol{\theta}_0) = E[g_1 (\partial/\partial\boldsymbol{\theta}^T)f_1(\boldsymbol{\theta}_0)],$$

Denoting

$$A = E\left[\left(\partial/\partial\boldsymbol{\theta}^T\right)f_1(\boldsymbol{\theta}_0)\left(\partial/\partial\boldsymbol{\theta}^T\right)f_1(\boldsymbol{\theta}_0)\right], \quad b = E\left[\left(\partial/\partial\boldsymbol{\theta}^T\right)f_1(\boldsymbol{\theta}_0)\right]g_1,$$

it follows now from (5) and Assumptions A-C that

$$\eta(\xi) = E[g_1 \phi_1(\xi)] = E\left[g_1 - b^T A^{-1} (\partial/\partial\boldsymbol{\theta}^T)f_1(\boldsymbol{\theta}_0) w_1(\xi)\right]$$

and

$$(12) \quad E\left[(\partial/\partial\boldsymbol{\theta}^T)f_1(\boldsymbol{\theta}_0)\phi_1(\xi)\right] \equiv 0.$$

However, assuming again that $\boldsymbol{\theta}_0$ is in the interior of Θ , it follows easily from the first-order

condition for a minimum of the sum of squared residuals that under the global alternative involved $b = 0$, and thus

$$(13) \quad \eta(\xi) = E[g_t w_t(\xi)].$$

It follows now from (13) that $\text{plim}_{n \rightarrow \infty} \hat{T}/n = \int \eta(\xi)^2 d\mu(\xi) > 0$ under all global alternatives for which $P[E(g_t | \mathfrak{S}_t) = 0] < 1$, provided the ICM test is potentially consistent. Note that this class of global alternatives may be smaller than the one for which $P(g_t = 0) < 1$. Only if the ICM test is consistent against all global alternatives of the latter type, the test is said to be consistent. Thus:

THEOREM 2: *Under Assumptions A-C the ICM test is consistent if and only if it is potentially consistent and $\mathfrak{S}_t = \mathfrak{G}_t$.*

Another way of looking at the notions of consistency and potential consistency is the following.

DEFINITION 2: *Let \mathfrak{L}_t be the Hilbert space (with the usual scalar product) spanned by $w_t(\xi)$ for all $\xi \in \Xi$ for which an arbitrarily small open neighborhood has positive μ measure.*

Then the ICM test is potentially consistent if and only if $\mathfrak{L}_t = L^2(\mathfrak{S}_t)$, where the latter is the set of all \mathfrak{S}_t measurable random variables with finite second moments. The ICM test is consistent if and only if $\mathfrak{L}_t = L^2(\mathfrak{G}_t)$.

Any g_t in $L^2(\mathfrak{G}_t)$ can be decomposed in $g_t = g_{1,t} + g_{2,t}$, where $g_{1,t}$ is the projection of g_t on \mathfrak{L}_t and $g_{2,t}$ is the residual. Then by construction, $E[g_{2,t} w_t(\xi)] = 0$ a.s. μ , hence $\eta(\xi) = E[g_{1,t} w_t(\xi)]$ a.s. μ . This implies that only the projection $g_{1,t}$ of g_t on \mathfrak{L}_t can be detected by the ICM test. Therefore, if the class of alternatives g_t is not completely contained in the Hilbert space

\mathfrak{G}_t , then there exist global alternatives $g_{2,t}$ for which the ICM test has asymptotic power less than 1, and local alternatives for which the test has only trivial power. The latter follows from Bierens and Ploberger (1994, Corollary 1), where we have shown that the ICM test has nontrivial local power if and only if $\int \eta(\xi)^2 d\mu(\xi) > 0$. The consistency condition $\mathfrak{G}_t = L^2(\mathfrak{G}_t)$ renders the existence of such alternatives impossible.

3.2. Orthogonal alternatives

Note that the result (13) does not necessarily hold under the local alternative (3). Thus under the local alternative and Assumption C we have

$$(14) \quad \eta(\xi) = E[v_t w_t(\xi)],$$

where

$$(15) \quad v_t = g_t - b^T A^{-1} (\partial/\partial \theta^T) f_t(\theta_0)$$

are the errors of the linear regression of g_t on $(\partial/\partial \theta^T) f_t(\theta_0)$. It follows now from (14) that the condition $\int \eta(\xi)^2 d\mu(\xi) > 0$ for nontrivial local power holds if and only if the ICM test is potentially consistent and $P(E[v_t | \mathfrak{F}_t] = 0) < 1$. Consequently, the ICM test has only trivial asymptotic power against local alternatives for which g_t is an exact linear combination of the components of $(\partial/\partial \theta^T) f_t(\theta_0)$, as then $v_t = 0$ a.s. Note however that for the global alternative model (11) such a g_t cannot occur, because it follows from $b = 0$ that if g_t is an exact linear combination of the components of $(\partial/\partial \theta^T) f_t(\theta_0)$ then $g_t = 0$.

Although the condition $b = 0$ is not strictly necessary for the consistency of the least squares estimator $\hat{\theta}$ under the local alternative (3), the fact that this condition must hold for global alternatives suggests that restricting the local alternatives to those for which $b = 0$ does not cause too much loss of generality. Therefore, in the sequel of this paper we shall confine our attention to *orthogonal* local alternatives only:

DEFINITION 3: *Local alternative of the type (3) are said to be orthogonal if $E[g_t (\partial/\partial \theta^T) f_t(\theta_0)] = 0$.*

THEOREM 3: *Let Assumptions A-C hold and let the ICM test be potentially consistent. Then the ICM test has nontrivial asymptotic power against all orthogonal local alternatives for which $P[E(g_t|\mathfrak{S}_t) = 0] < 1$. Consequently, if the ICM test is consistent then its local asymptotic power is nontrivial, and if the asymptotic power of the ICM test against all orthogonal local alternatives is nontrivial then the ICM test is consistent.*

4. LARGE LOCAL ALTERNATIVES

Consider the following class of "large" *orthogonal* local alternatives:

$$(16) \quad \begin{aligned} H_1^L(c): \quad & y_{t,n} = f_t(\theta_0) + c\sigma g_t/\sqrt{n} + u_t, \quad (t = 1, \dots, n), \text{ with} \\ & E(g_t^2) = 1, \quad E[g_t(\partial/\partial\theta^T)f_t(\theta_0)] = 0 \text{ and } E(u_t^2|\mathfrak{A}) = \sigma^2 \text{ a.s.,} \end{aligned}$$

where c is a "large" positive constant. Clearly, the standardization $E(g_t^2) = 1$ does not cause any loss of generality, and the same applies to the factor σ in front of g_t . The present form of the large local alternative involved has been chosen for convenience. The results below can be straightforwardly adjusted for the case where c in (16) is replaced by c/σ and the condition $E(g_t^2) = 1$ is replaced by $0 < E(g_t^2) < \infty$.

THEOREM 4: *Let Assumptions A-C hold and denote the asymptotic power function of the ICM test by $\Pi_1(c) = \lim_{n \rightarrow \infty} P(\text{ICM-test rejects } H_0 | H_1^L(c))$. If the ICM test is potentially consistent, then*

$$\lim_{c \rightarrow \infty} \frac{\ln[1 - \Pi_1(c)]}{c^2} = -\frac{1}{2}(E(g_1|\mathfrak{S}_1))^2 \geq -\frac{1}{2}.$$

If the ICM test is either consistent, or potentially consistent and $E(g_1|\mathfrak{S}_1) = g_1$, or if merely $g_1 \in \mathfrak{A}_1$, then

$$\lim_{c \rightarrow \infty} \frac{\ln[1 - \Pi_1(c)]}{c^2} = -\frac{1}{2}.$$

PROOF: Appendix

Note that the latter result implies that for each δ in the interval (0,1) we can find a c_δ such that for all $c > c_\delta$,

$$(17) \quad \exp\left[-\frac{1}{2}(1 + \delta)c^2\right] \leq 1 - \Pi_1(c) \leq \exp\left[-\frac{1}{2}(1 - \delta)c^2\right].$$

Thus, for $c \rightarrow \infty$ the asymptotic power function of the consistent ICM-test approaches 1 at an exponential rate. Moreover, note that the result in Theorem 4 for the consistent ICM-test is remarkable in that it neither depends on the choice of the weight function w_t and the probability measure μ , nor on the significance level, as long as these choices preserve consistency.

The result in Theorem 4 is even more remarkable if we compare it with the asymptotic t -test of the null hypothesis $\delta_0 = 0$ in the auxiliary regression $y_t = f_t(\theta_0) + \delta_0 g_t^* + u_t$, where g_t^* is some "guess" of the g_t in (16), which for the sake of a fair comparison is assumed to satisfies the same conditions as for g_t , i.e., g_t^* is measurable w.r.t. \mathfrak{G}_t , $E[(g_t^*)^2] = 1, E[g_t^*(\partial/\partial\theta^T)f_t(\theta_0)] = 0$. Denoting the least squares estimator of δ_0 by $\hat{\delta}$, it is a standard exercise to show that under Assumptions A-C and the local alternative (16), $\sqrt{n}\hat{\delta} \rightarrow N(c\sigma\rho, \sigma^2)$ distr., where $\rho = \text{corr}(g_t, g_t^*)$. Therefore, under the local alternative (16) the t -statistic $\hat{t}_\delta(c)$ of $\hat{\delta}$ satisfies $\hat{t}_\delta(c) \rightarrow N(c\rho, 1)$ in distr. Similarly to the proof of Theorem 4 (in the appendix) it is now easy to show:

THEOREM 5: *Let Assumptions A-C hold, and denote the asymptotic power function of the t -test by*

$$\Pi_2(c) = \lim_{n \rightarrow \infty} P(t\text{-test rejects } H_0 | H_1^{L(c)}).$$

Then

$$\lim_{c \rightarrow \infty} \frac{\ln[1 - \Pi_2(c)]}{c^2} = -\frac{1}{2}\rho^2,$$

where ρ is the correlation coefficient of g_t^ and g_t .*

Similarly to (17) this result implies that for each δ in the interval $(0,1)$ we can find a c_δ such that for all $c > c_\delta$,

$$(18) \quad \exp\left[-\frac{1}{2}(1 + \delta)c^2\rho^2\right] \leq 1 - \Pi_2(c) \leq \exp\left[-\frac{1}{2}(1 - \delta)c^2\rho^2\right].$$

Comparing (17) and (18) we see that if the correlation coefficient ρ involved is *not* equal to -1 or $+1$, then there exists a c_0 such that $\Pi_1(c) > \Pi_2(c)$ for $c > c_0$. Thus the asymptotic power function of the consistent ICM test converges to 1 at the same rate as the asymptotic power function of the t -test only if $g_t^* = g_t$. Theorem 5, though, implies that the t -test is consistent against all global alternatives for which g_t^* and g_t have nonzero correlation, but as long as the correlation between g_t^* and g_t is not perfect the ICM test is more powerful than the t -test, uniformly for large c 's.

APPENDIX

PROOF OF THEOREM 4: Theorem 4 follows straightforwardly from Lemmas A.1-4 below:

LEMMA A.1: *Let K be an arbitrary positive constant. Under Assumption A and the local alternative (16),*

$$(A1) \quad \lim_{c \rightarrow \infty} \frac{\ln(P(T(c) \leq K))}{c^2} = -\frac{1}{2}\sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i.$$

PROOF: First, observe from Theorem 1 that the limiting distribution of the ICM test statistic under this "large" local alternative is:

$$T(c) = \sum_{i=1}^{\infty} (c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2, \quad \varepsilon_i \text{ is i.i.d. } N(0,1).$$

Next, observe that for $x > 0$,

$$\int_x^{\infty} \exp[-\frac{1}{2}u^2] du \leq \int_x^{\infty} \frac{u}{x} \exp[-\frac{1}{2}u^2] du = \frac{\exp[-\frac{1}{2}x^2]}{x}$$

and

$$\int_x^{\infty} \exp[-\frac{1}{2}u^2] du \geq e^{1/2} \int_x^{\infty} \frac{u}{x} \exp[-\frac{1}{2}u^2/x^2] \exp[-\frac{1}{2}u^2] du = \frac{\exp[-\frac{1}{2}x^2]}{x(1+1/x^2)}$$

hence

$$(A2) \quad 1 - \Phi(x) \sim \frac{\exp(-x^2/2)}{x\sqrt{2\pi}} \text{ for } x \rightarrow \infty.$$

Consequently, we have

$$(A3) \quad \lim_{x \rightarrow \infty} \frac{\ln(\Phi(-x))}{x^2} = \lim_{x \rightarrow \infty} \frac{\ln(1-\Phi(x))}{x^2} = -\frac{1}{2},$$

where Φ is the c.d.f. of the standard normal distribution. The result (A2) implies that for every

constant M ,

$$\frac{\Phi(M-x)}{\Phi(-x)\exp(Mx)} \sim \exp(-\frac{1}{2}M^2) \text{ for } x \rightarrow \infty$$

hence, using (A3), it follows that for every $M > 0$,

$$\lim_{|x| \rightarrow \infty} \frac{\ln(\Phi(M-x) - \Phi(-M-x))}{x^2} = -\frac{1}{2}.$$

In its turn this result implies that for $M_i > 0$,

$$(A4) \quad \lim_{c \rightarrow \infty} \frac{\ln \left(P((c\eta_i + \varepsilon_i \sqrt{\lambda_i})^2 \leq M_i) \right)}{c^2} = -\frac{1}{2} \eta_i^2 / \lambda_i.$$

Note that this result also holds if $\eta_i^2 / \lambda_i = 0$.

The result (A4) now enables us to prove the lemma in two steps. First, we establish the upperbound of the limit (A1), and then the lowerbound.

Step 1: For every $K > 0$ and every natural number $N > 1$ we have

$$\begin{aligned} P(T(c) \leq K) &\leq P\left(\bigcap_{i=1}^N \{(c\sigma\eta_i + \varepsilon_i \sqrt{\lambda_i})^2 \leq K\}\right) \\ &= \prod_{i=1}^N P((c\sigma\eta_i + \varepsilon_i \sqrt{\lambda_i})^2 \leq K), \end{aligned}$$

hence it follows from (A4) that for arbitrary $N > 1$,

$$(A5) \quad \limsup_{c \rightarrow \infty} \frac{\ln[P(T(c) \leq K)]}{c^2} \leq -\frac{1}{2} \sigma^2 \sum_{i=1}^N \eta_i^2 / \lambda_i.$$

Letting $N \rightarrow \infty$, (A5) implies:

$$(A6) \quad \limsup_{c \rightarrow \infty} \frac{\ln[P(T(c) \leq K)]}{c^2} \leq -\frac{1}{2} \sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i.$$

Step 2: For arbitrary $K > 0$ and natural numbers $N > 1$ we have

$$\begin{aligned}
& P(T(c) \leq K) \\
& \geq P\left(\bigcap_{i=1}^N \{(c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2 \leq \frac{K}{2N}\} \cap \left\{\sum_{i=N+1}^{\infty} (c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2 \leq \frac{K}{2}\right\}\right) \\
& = \left(\prod_{i=1}^N P[(c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2 \leq \frac{K}{2N}]\right) P\left(\sum_{i=N+1}^{\infty} (c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2 \leq \frac{K}{2}\right).
\end{aligned}
\tag{A7}$$

Moreover, for any fixed $L > N$ we have

$$\begin{aligned}
& P\left(\sum_{i=N+1}^L (c\sigma\eta_i + \varepsilon_i\sqrt{\lambda_i})^2 \leq \frac{K}{2}\right) \\
& = \int I\left(\sum_{i=N+1}^L \lambda_i x_i^2 \leq \frac{K}{2}\right) \prod_{i=N+1}^L \left(\frac{\exp\left(-\frac{1}{2}(x_i - c\sigma\eta_i/\sqrt{\lambda_i})^2\right)}{\sqrt{2\pi}} dx_i\right) \\
& = \int I\left(\sum_{i=N+1}^L \lambda_i x_i^2 \leq \frac{K}{2}\right) \frac{\exp\left(-\frac{1}{2}\sum_{i=N+1}^L x_i^2\right)}{(\sqrt{2\pi})^{L-N}} \exp\left(c\sigma \sum_{i=N+1}^L x_i \eta_i/\sqrt{\lambda_i}\right) \\
& \quad \times \exp\left(-\frac{1}{2}c^2\sigma^2 \sum_{i=N+1}^L \eta_i^2/\lambda_i\right) dx_{N+1} \dots dx_L \\
& \geq \exp\left(-\frac{1}{2}c^2\sigma^2 \sum_{i=N+1}^L \eta_i^2/\lambda_i\right) \\
& \quad \times P\left(\sum_{i=N+1}^L \lambda_i \varepsilon_i^2 \leq \frac{K}{2} \wedge \sum_{i=N+1}^L \varepsilon_i \sigma \eta_i/\sqrt{\lambda_i} \geq 0\right),
\end{aligned}$$

hence, letting $L = \infty$, it follows that

$$\begin{aligned}
& \liminf_{c \rightarrow \infty} \frac{\ln \left[P \left(\sum_{i=N+1}^{\infty} (c \sigma \eta_i + \varepsilon_i \sqrt{\lambda_i})^2 \leq \frac{K}{2} \right) \right]}{c^2} \geq -\frac{1}{2} \sigma^2 \sum_{i=N+1}^{\infty} \eta_i^2 / \lambda_i \\
& + \liminf_{c \rightarrow \infty} \frac{\ln \left[P \left(\sum_{i=N+1}^{\infty} \lambda_i \varepsilon_i^2 \leq \frac{K}{2} \wedge \sum_{i=N+1}^{\infty} \varepsilon_i \sigma \eta_i / \sqrt{\lambda_i} \geq 0 \right) \right]}{c^2} \\
& = -\frac{1}{2} \sigma^2 \sum_{i=N+1}^{\infty} \eta_i^2 / \lambda_i,
\end{aligned} \tag{A8}$$

where the last conclusion follows from the fact that the log of the probability at the right-hand side of (A8) is finite and does not depend on c . Combining (A4), (A7) and (A8) now yields

$$\liminf_{c \rightarrow \infty} \frac{\ln[P(T(c) \leq K)]}{c^2} \geq -\frac{1}{2} \sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i. \tag{A9}$$

The theorem under review now follows from (A6) and (A9).

LEMMA A.2: *Let Assumptions A-C hold. If the ICM test is consistent then under the local alternative (16), $\sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i = 1$.*

PROOF: From Assumption A and the conditions in (16), it follows that $E[\phi_t(\xi_1) \phi_t(\xi_2)] = \Gamma(\xi_1, \xi_2) / \sigma^2$. Moreover, part (10) of Theorem 1, together with the conditions of the lemma under review, implies that $E[g_t \int \phi_t(\xi) \psi_t(\xi) d\mu(\xi)] = \eta_t$. Now let β_t be a sequence of coefficients. Then it follows from (12) and (15) that for each t we have

$$\begin{aligned}
& E \left(g_t - \sum_{i=1}^{\infty} \beta_i \int_{\Xi} \phi_i(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 = 1 - 2 \sum_{i=1}^{\infty} \beta_i \int_{\Xi} \eta_i(\xi) \psi_i(\xi) d\mu(\xi) \\
& + \sigma^2 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \beta_i \beta_j \iint_{\Xi \times \Xi} \Gamma(\xi_1, \xi_2) \psi_i(\xi_1) \psi_j(\xi_2) d\mu(\xi_1) d\mu(\xi_2) \\
& = 1 - 2 \sum_{i=1}^{\infty} \beta_i \eta_i + \sum_{i=1}^{\infty} \beta_i^2 \lambda_i / \sigma^2 \geq 1 - \sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i \geq 0,
\end{aligned}
\tag{A10}$$

for $\beta_i = \sigma^2 \eta_i / \lambda_i$. Note that $\lambda_i = 0$ implies $\eta_i = 0$, as otherwise we can choose β_i such that the left-hand side of (A10) becomes negative. Therefore we may assume that $\eta_i^2 / \lambda_i = 0$ if $\lambda_i = 0$.

Finally, we show that the expectation at the left-hand side of (A10) is zero. A necessary and sufficient condition for this is that:

$$g_t - \sigma^2 \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \phi_i(\xi) \psi_i(\xi) d\mu(\xi) = 0 \text{ a.s.}
\tag{A11}$$

By Theorem 2 the consistency of the ICM test implies potential consistency and $g_t = E(g_t | \mathfrak{F}_t)$ a.s. Therefore, taking a_t in Definition 1 to be the random variable at the left-hand side of (A11), it follows that (A11) is true if

$$\int \left[E \left(g_t - \sigma^2 \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \phi_i(\xi) \psi_i(\xi) d\mu(\xi) \right) w_t(\xi_*) \right] d\mu(\xi_*) = 0.
\tag{A12}$$

Since the local alternative under review is orthogonal, we may replace $w_t(\xi_*)$ by $\phi_t(\xi_*)$. Then

$$\begin{aligned}
& \int_{\Xi} \left[E \left(g_t - \sigma^2 \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \phi_t(\xi) \psi_i(\xi) d\mu(\xi) \right) \phi_t(\xi_*) \right] d\mu(\xi_*) \\
&= \int_{\Xi} \left(E[g_t w_t(\xi_*)] - \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \sigma^2 E[\phi_t(\xi) \phi_t(\xi_*)] \psi_i(\xi) d\mu(\xi) \right) d\mu(\xi_*) \\
&= \int_{\Xi} \left[\eta(\xi_*) - \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \Gamma(\xi_*, \xi) \psi_i(\xi) d\mu(\xi) \right] d\mu(\xi_*) \\
&= \int_{\Xi} \left[\eta(\xi) - \sum_{i=1}^{\infty} \eta_i \psi_i(\xi) \right] d\mu(\xi),
\end{aligned}$$

where the second equality follows from (7), (9), (13) and the third equality from (8), it follows now that (A11) is true if $\eta(\xi) = \sum_{i=1}^{\infty} \eta_i \psi_i(\xi)$ a.s. μ . The latter is true because $\{\psi_i(\cdot)\}$ is a complete orthonormal basis of the space $L^2(\mu)$, so that any function η in $L^2(\mu)$ can be written (a.s. μ) as a linear combination of the $\psi_i(\xi)$'s, with (Fourier) coefficients given by (10).

LEMMA A.3: *Under the conditions of Lemma A.2, except that now the ICM test is only potentially consistent, we have $\sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i = E(E(g_t | \mathfrak{I}_t))^2 \leq 1$.*

PROOF: Inspection of the proof of Lemma 2 reveals that without the consistency condition equation (A10) becomes

$$\begin{aligned}
& E \left(E(g_t | \mathfrak{I}_t) - \sigma^2 \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int_{\Xi} \phi_t(\xi) \psi_i(\xi) d\mu(\xi) \right)^2 \\
&= E(E(g_t | \mathfrak{I}_t))^2 - \sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i \geq 0.
\end{aligned}$$

Moreover, $1 = E(g_t^2) = E([g_t - E(g_t | \mathfrak{I}_t)]^2) + E([E(g_t | \mathfrak{I}_t)]^2)$. Therefore, the lemma is easy to verify from the proof of Lemma A.2.

LEMMA A.4: Let Assumptions A-C hold. If $g_1 \in \mathfrak{L}_1$ then under the local alternative (16), $\sigma^2 \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i = 1$.

PROOF: Under the condition of the lemma the left hand side of (A11) is an element of \mathfrak{L}_1 , i.e., it is a linear combination of $w_i(\xi)$'s, hence (A12) implies that

$$E \left[\left(g_1 - \sigma^2 \sum_{i=1}^{\infty} (\eta_i / \lambda_i) \int \phi_i(\xi) \psi_i(\xi) d\mu(\xi) \right) w_i(\xi) \right] = 0 \text{ a.s.}$$

for all $w_i(\xi)$'s in this linear combination, which in its turn implies that (A11) holds.

REFERENCES:

- BIERENS, H.J. (1982): "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134
- BIERENS, H.J. (1984): "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353
- BIERENS, H.J. (1987): "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics* 35, 161-190
- BIERENS, H.J. (1988): "ARMA Memory Index Modeling of Economic Time Series (with discussion)", *Econometric Theory* 4, 35-59
- BIERENS, H.J. (1990): "A Consistent Conditional Moment Test of Functional Form", *Econometrica* 58, 1443-1458
- BIERENS, H.J. and J. HARTOG (1988): "Nonlinear Regression with Discrete Explanatory Variables, with an Application to the Earnings Function", *Journal of Econometrics* 38, 269-299
- BIERENS, H.J. and W. PLOBERGER (1994): "Asymptotic Theory of Integrated Conditional Moment Tests", Working Paper, Department of Economics, Southern Methodist University, Dallas
- DE JONG, R.M. (1995): "On the Bierens Test Under Data Dependence", *Journal of Econometrics* (forthcoming) Amsterdam

DE JONG,R.M. and H.J.BIERENS (1994): "On the Limiting Behavior of a Chi-Square Type Test If the Number of Conditional Moments Tested Approaches Infinity", *Econometric Theory* 9, 70-90

FAN,Y. and Q.LI (1992a), "A General Nonparametric Model Specification Test", Mimeo, Department of Economics, University of Windsor, Ontario

FAN,Y. and Q.LI (1992b), "The Asymptotic Expansion of Kernel Sum of Squared Residuals and Its Application in Hypothesis Testing", Mimeo, Department of Economics, University of Windsor, Ontario

FAN,Y. and Q.LI (1992c), "A Consistent Nonparametric Hausman Test", Mimeo, Department of Economics, University of Windsor, Ontario

GOZALO,P.L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models", *Econometric Theory* 9, 451-477

HANSEN,B.E. (1990): "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis", Rochester Center for Economic Research Working Paper 296, University of Rochester.

HARDLE,W. and E.MAMMEN (1993): "Comparing Nonparametric Versus Parametric Regression Fits", *Annals of Statistics* 21, 1926-1947

HONG,Y. and H.WHITE (1991): "Consistent Specification Testing Via Nonparametric Series Regression", Mimeo, Department of Economics, University of California, San Diego.

HOROWITZ,J.L. and W.HARDLE (1992): "Testing a Parametric Model Against a Semiparametric Alternative", Mimeo, Department of Economics, University of Iowa.

LEWBEL,A. (1992): "Consistent Nonparametric Hypothesis Testing with an Application to Slutsky Symmetry", Mimeo, Department of Economics, Brandeis University.

STINCHCOMBE,M.B. and H.WHITE (1991): "Consistent Specification Tests Using Duality", Mimeo, Department of Economics, University of California, San Diego.

YATCHEW,A.J. (1992): "Nonparametric Regression Tests Based on Least Squares", *Econometric Theory* 8, 435-451

WHITE,H. (1989): "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks", *Proceedings of the International Joint Conference on Neural Networks, Washington, D.C.*, San Diego: SOS Printing, II: 451-455

WHITE,H. and Y.HONG (1993): "M-Testing Using Finite and Infinite Dimensional Parameter Estimators", Mimeo, Department of Economics, University of California, San Diego

WOOLDRIDGE,J.M. (1992): "A Test for Functional Form Against Nonparametric Alternatives", *Econometric Theory* 8, 452-475

ZHENG,J.X. (1993): "A Consistent Test for Goodness of Fit of Parameter Distribution", Mimeo, Department of Economics, University of Texas at Austin